FINAL TECHNICAL REPORT

PhaseNet: Towards Comprehensive P and S Wave Arrival Time Measurements Using Machine Learning

Gregory C. Beroza

Stanford University
Department of Geophysics
397 Panama Mall
Stanford, CA, 94305-2215
Telephone: (650)723-4746; fax: (650) 725-7344

Principal Investigator: Gregory C. Beroza (beroza@stanford.edu)
Keywords: Machine Learning, Earthquake Monitoring
Program Element III: U. S. Geological Survey National Earthquake Hazards Reduction Program
Award Number G19AP00078.
Project Period: 7/01/2019-6/30/2020

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number G20AP00015. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey.

Abstract

As the number of seismic sensors grows, it is increasingly difficult to pick seismic phases manually and comprehensively, yet such efforts are fundamental to earthquake monitoring. Despite years of improvements in automatic phase picking, it has been difficult to match the performance of experienced analysts. Different seismic analysts also pick phases differently, which introduces bias into earthquake monitoring. Under this grant we developed a deep-neural-network-based arrival-time picking method called "PhaseNet" that picks the arrival times of both *P* and *S* waves. PhaseNet uses either single-component or three-component seismic waveforms as input and generates continuous probability distributions of *P* arrivals, *S* arrivals and noise as output. We train PhaseNet such that peaks in the probability distributions provide accurate arrival times for both *P* and *S* waves. PhaseNet is trained on a waveform data set assembled from analyst-labelled *P* and *S* arrival times from the Northern California Earthquake Data Center. The data set includes over 700,000 waveform samples from over three decades of earthquake recordings. PhaseNet achieves much better picking accuracy and recall than existing methods when applied to the waveforms of known earthquakes. In subsequent studies using PhaseNet to pick arrival times in continuous data, we have realized its potential to increase dramatically the number of *S*-wave observations over what would otherwise be available. Our experiments indicate that with PhaseNet it is actually easier to pick *S* waves for small earthquakes than it is to pick *P* waves because the *S* waves have larger amplitude. This new capability will enable both improved locations and improved shear wave velocity models.

Introduction

Deep neural networks learn the features from labelled data, both noise and signal, which proves a powerful advantage for complex seismic waveforms. Our PhaseNet network is trained on a catalog of available *P* and *S* arrival times picked by experienced analysts. Unfiltered three-component seismic waveforms are the input to PhaseNet, which is trained to output three probability distributions: *P* wave, *S* wave and noise. Peaks in the *P* wave and *S* wave probability are designed to correspond to the predicted *P* and *S* arrival times. PhaseNet provides high accuracy and recall rate for both *P* and *S* picks, and achieves significant improvement compared with a traditional characteristic-function-based methods. PhaseNet has the potential to provide comprehensive, superior performance for standard earthquake monitoring.

Data

We gathered digital seismic waveform data based on the Northern California Earthquake Data Center Catalog (NCEDC 2014). We use three-component data that have both *P* and *S* arrival times. This yielded 779,514 recordings. We use stratified sampling based on stations to divide this data set into training, validation and test data sets, with 623,054, 77,866 and 78,592 samples, respectively. The training and validation sets are used during training, fine-tuning parameters and model selection. The test set is only used to evaluate the final performance. This data set has a diversity of waveform characteristics. It includes a variety of instruments in the Northern California Seismic Network and covers a wide range of signal-to- noise ratio (SNR). The complexity of this data set makes it challenging for automatic phase picking, but it is representative of the data available.

We apply minimal data pre-processing to the data. We randomly select a 30-s time window that includes the *P* and *S* arrival times as the input of PhaseNet. The position of the arrivals within the window are varied to ensure that the algorithm does not just learn the windowing scheme or the positioning within it. All data are sampled at 100 Hz, which is the most common sampling rate in the raw data set, so that the 30-s input waveforms have 3001 data points for each component. We normalize each component waveform by removing its mean and dividing it by the standard deviation. There are errors in the data such that manually picked time points in the data set may not be the true *P*/*S* arrivals, but we expect the ground-truth arrival times will be centered on the manual picks. For this reason, we apply a mask with the shape of a Gaussian distribution around the manual picks. The time point picked by analysts is assigned the highest probability, while the nearby data points have reduced probabilities (Fig. 4d). The standard deviation of the Gaussian distribution is set to 0.1 s in each case. Representing manual picks probabilistically should allow the algorithm to reduce the influence of picking errors in the data set. Because we have considered the probabilities of the nearby data points, the mask increases the amount of information in *P* and *S* picks relative to noise and helps accelerate convergence. Here the noise includes all data points that are not first arrivals of *P* or *S* waves.

Method

The architecture of PhaseNet (Fig. 1) is modified from U-Net (*Ronneberger et al.* 2015) to deal with 1-D time-series data. U-net is a deep neural network approach used in biomedical image processing that seeks to localize properties in an image through an encoding-decoding structure.

The mapping to our problem is to localize the properties of our time-series into three classes: *P* pick, *S* pick and noise. The inputs are three-component seismograms of known earthquakes. The outputs are probability distributions of *P* wave, *S* wave and noise. In our experiments, the input and output sequences contain 3001 data points for each component (30 s long, sampled at 100 Hz).

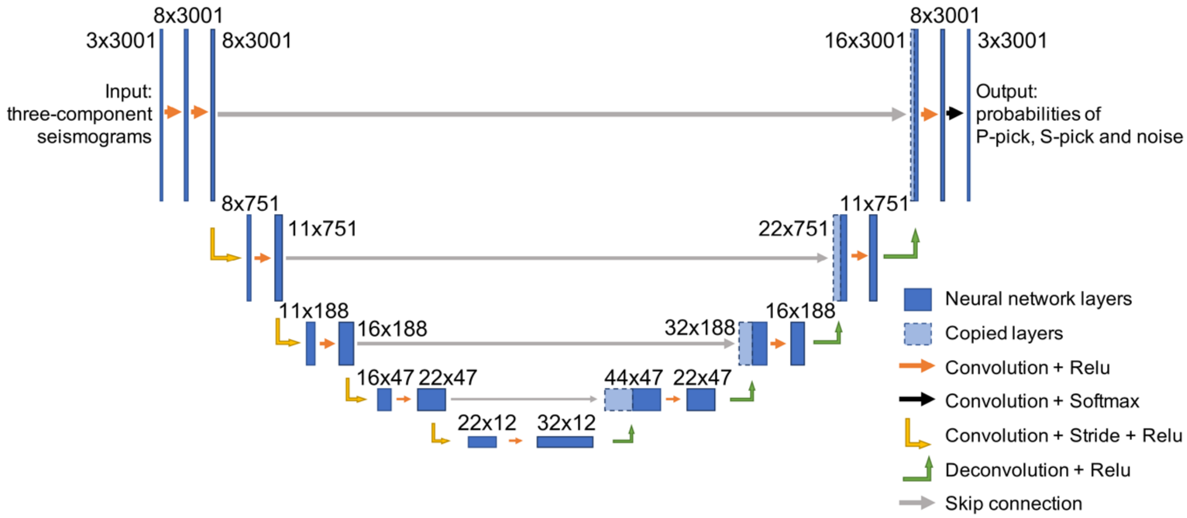
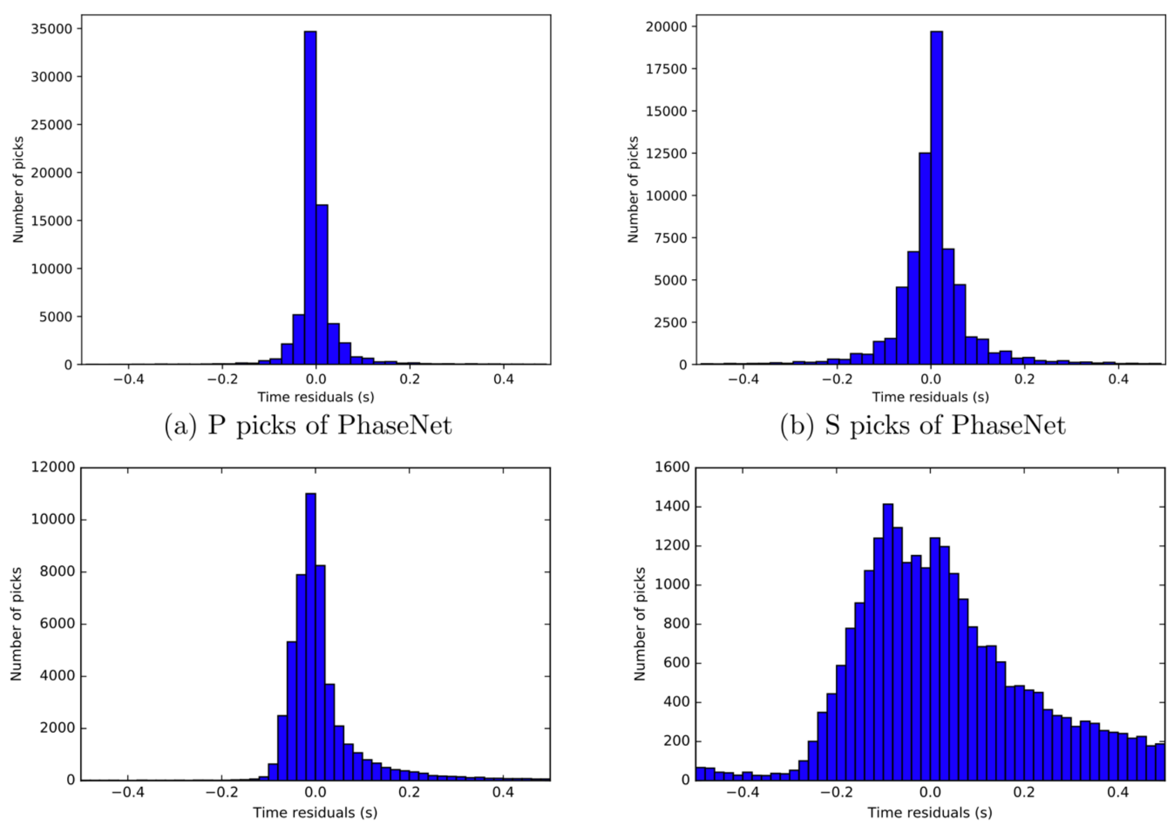


Fig. 1. PhaseNet architecture. Input are 30-s three-component seismograms sampled at 100 Hz. Output is three probabilities with the same length as input for *P*, *S,* and noise. Blue rectangles represent layers inside the neural network. The numbers near them are the dimensions of each layer, which follow a format of "number of channels × length of each channel". Arrows are operations applied between layers, as noted in the low right corner. Input seismic data go through four down-sampling stages and four up-sampling stages. The down-sampling is done by 1-D convolution and stride. We set the length of convolution kernel to seven data points and the stride step to four data points. Up-sampling is done by deconvolution, which recovers the input length of the previous stage. A skip connection at each stage concatenates the left output to the right layer without going through the deeper layers, which improves convergence during training. Blue rectangles with dashed boundaries are layers copied directly by the skip connection. The softmax normalized exponential function sets probabilities in the last layer.

The input seismic data go through four down-sampling stages and four up-sampling stages. Inside each stage, we apply 1-D convolutions and rectified linear unit (ReLU) activations. The down-sampling process is designed to extract and shrink the useful information from raw seismic data to a few neurons, so each neuron in the last layer makes up a broadly receptive window. The up-sampling process expands and converts this information into probability distributions of P wave, S wave and noise for each time point. A skip connection at each depth directly concatenates the left output to the right layer without going through the deeper layer. This helps improve convergence during training (*Ronneberger et al.* 2015; *Li et al.* 2017). The 1-D convolution size is set to seven data points. The stride step for downsampling is set to four data points, so after each stride the channel length is condensed into one-fourth of its original dimension, while the deconvolution operation for up-sampling expands the condensed layers by a factor of four to recover its previous length. We have added padding at the front and the back of each layer during convolutions to make the input and output sequences have the same length. Fig. 1 shows the size of each layer and the operations of convolution and deconvolution. The softmax normalized exponential function is used to set probabilities in the last layer.

Results

We compare our results with those obtained by the open-source "AR picker" (*Akazawa* 2004) implemented in Obspy (*Beyreuther et al.* 2010). The results of both PhaseNet and AR picker are shown in Table 1. For our data set, our method achieved significant improvement, particularly for *S* waves. Because *S* waves emerge from the scattered waves of the *P* coda, picking *S* arrivals is usually more challenging for automatic methods. Fig. 2 shows the distribution of time residuals between the automated and human-labelled *P* and *S* picks and Fig. 3 shows examples of picks. The residual distributions of the *P* picks are much narrower than for the *S* picks, which is consistent with the fact that *P* wave arrivals are expected to be clearer and hence easier to pick. The residual distributions of both *P* and *S* picks for PhaseNet are distinctly narrower and do not have obvious skew towards late estimates of AR picker.



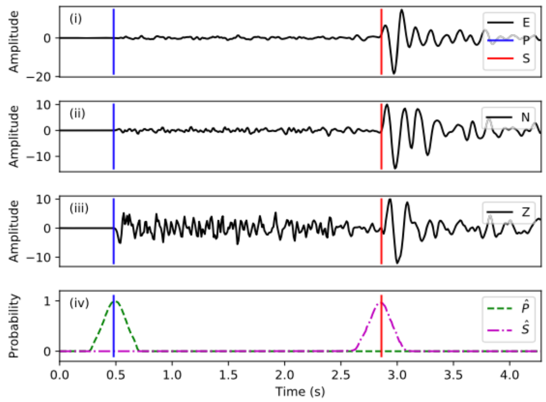
(a) P picks of PhaseNet

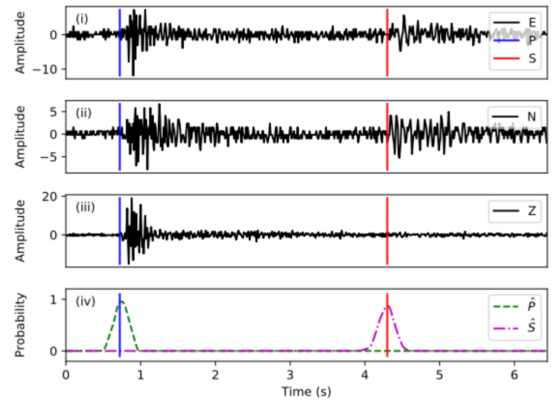
(b) S picks of PhaseNet

(c) P picks of AR picker

(d) S picks of AR picker

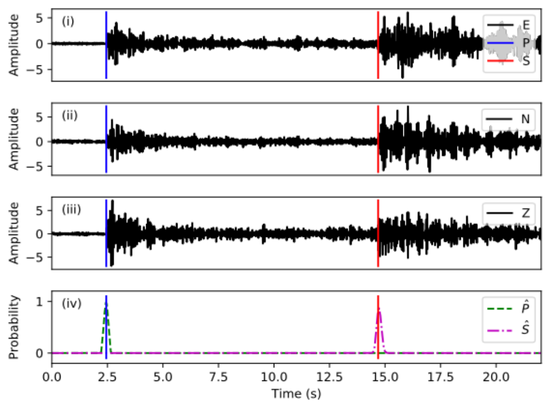
Fig. 2. The distribution of residuals of PhaseNet (upper panels) and AR picker (lower panels) on the test data set.

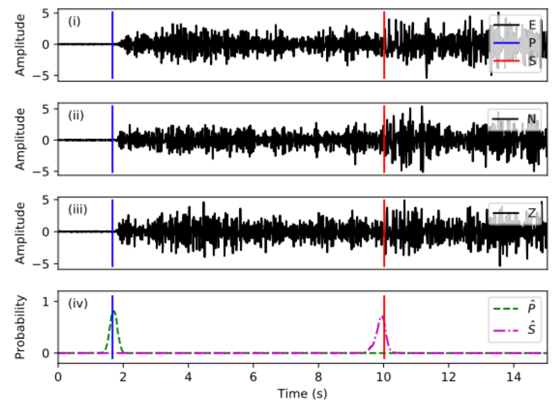




(a)

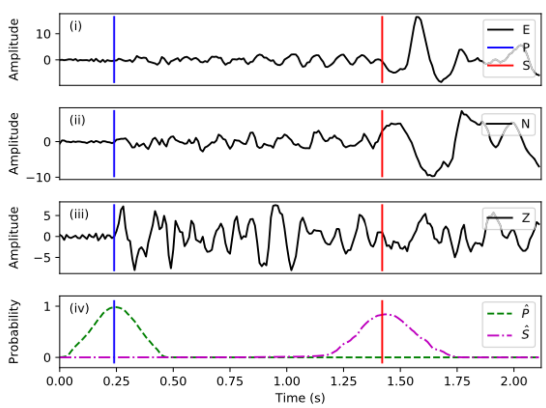
(b)





(c)

(d)



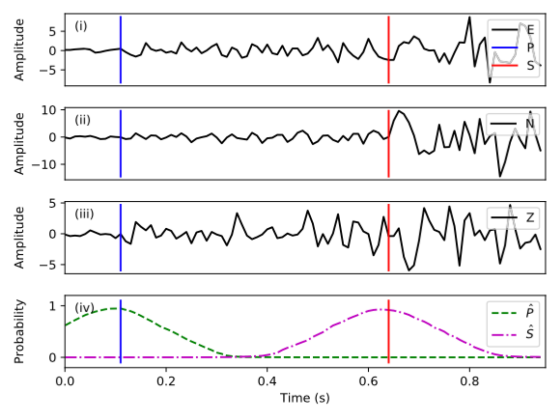


Fig 3. Six examples of PhaseNet output for the test data set showing its ability to pick arrival times reliably. The upper panels (i–iii) are the East-North-Vertical components of seismograms. The lower panels (iv) are the predicted probability distributions of *P* wave ($\hat{P}$) and *S* wave ($\hat{S}$). Blue and red lines are the *P* and *S* arrival times picked by analysts from the NCEDC phase catalog.

PhaseNet predicts the probability distributions of *P* and *S* picks for every data point in the time-series, so it cacn be applied to continuous data for earthquake detection. Figure 4 shows semi-synthetic continuous seismic data that we created by stacking waveforms of eight different events. These events are shifted to make the arrival-time interval between adjacent events equal to 6 s. We have applied both basic STA/LTA in Obspy and our PhaseNet method to this sequence. The short and long sliding windows of the STA/LTA method are set to 0.2 and 2 s, respectively. The output shows that PhaseNet produces similar spikes as STA/LTA methods; however, unlike these methods, PhaseNet also differentiates between *P* and *S* arrivals. This information can be used to reduce false detections and mis-associations.

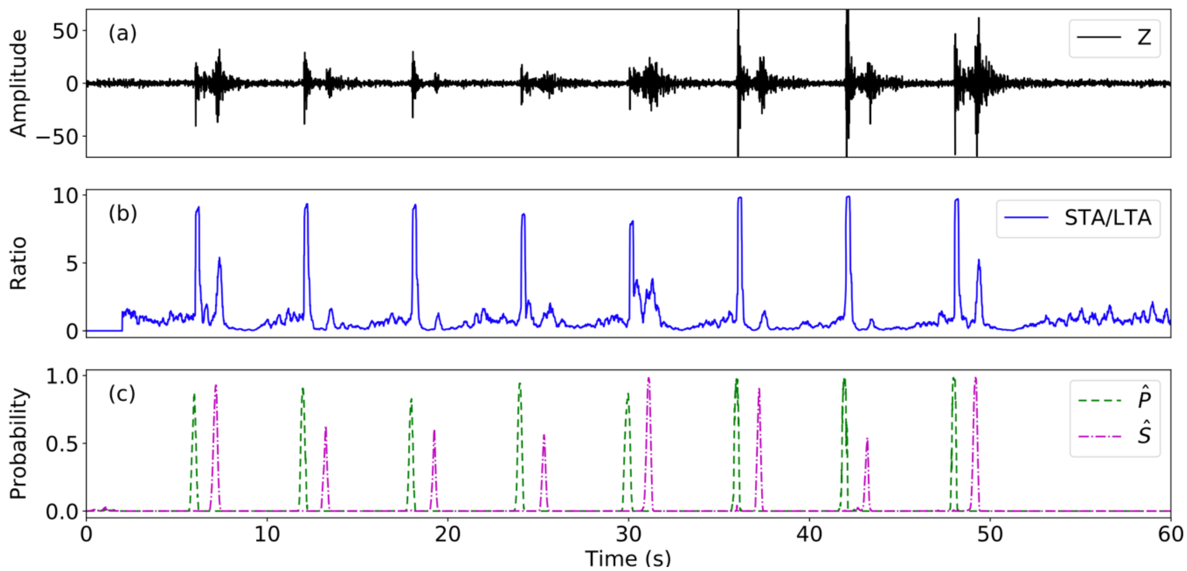


Fig. 4. Semi-synthetic continuous seismic waveforms. (a) Waveform of vertical component. (b) Output of basic STA/LTA in Obspy. (c) Output of PhaseNet. The continuous data are created by stacking waveforms of eight events. The first-arrival-time interval between adjacent events is 6 s. PhaseNet greatly outperforms STA/LTA in picking *S* waves and, unlike STA/LTA, actually discriminates between *P* and *S* picks.

Conclusions

Deep learning methods continue to rapidly. An important ingredient for improving them is the existence of large, labelled data sets. Seismology is fortunate to have such large data sets ready at hand in the form of decades of arrival times with accompanying waveforms. Under this grant we built a training data set using manually picked *P* and *S* arrival times from the Northern California Seismic Network catalog. Using that as input we developed PhaseNet, a deep neural network algorithm that uses three component waveform data to extract arrival times for *P* and *S* waves. Our method achieves significant improvements compared with existing methods, particularly for *S* waves. We have applied the PhaseNet method to earthquakes in California, Italy, China, Martinique, Mayotte, Kansas, and Oklahoma with great success. It will form the basis for an enhanced northern California earthquake catalog in the near future. The PhaseNet method was published [*Zhu and Beroza*, 2019] and the trained model is freely available on GitHub.

References

Akazawa, T. (2004) Technique for automatic detection of onset time of *P*- and *S*-phases in strong motion recordsin *13th World Conference on Earthquake Engineering,* Vancouver, Vol. **786,** pp. 786.

Baer, M. & Kradolfer, U. (1987) An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.,* **77**(4), 1437–1445.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. (2010) ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.,* **81**(3), 530–533.

Li, H., Xu, Z., Taylor, G., Studer, C. & Goldstein, T., 2017. Visualizing the loss landscape of neural nets, arXiv preprint arXiv:1712.09913.

Ronneberger, O., Fischer, P., and Brox, T. (2015) U-Net: convolutional networks for biomedical image segmentation, *Miccai,* **9351,** 234–241.

Zhu, W. and G. C. Beroza (2019) PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method, *Geophys. J. Int.,* **216** *261–273,* doi:10.1093/gji/ggy423.